Control of Buffer and Energy of a Wireless Device: Closed and Open Loop Approaches

V. S. Borkar, A. A. Kherani, Balakrishna Prabhu

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Control of Buffer and Energy of a Wireless Device: Closed and Open Loop Approaches

V. S. Borkar — A. A. Kherani — B. J. Prabhu

N° 5414

December 2004

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5414--FR+ENG

Control of Buffer and Energy of a Wireless Device: Closed and Open Loop Approaches

V. S. Borkar , A. A. Kherani , B. J. Prabhu *

Thème COM — Systèmes communicants
Projets Maestro

Rapport de recherche n° 5414 — December 2004 — 24 pages

Abstract: We consider a decision problem faced by an energy limited wireless device that operates in discrete time. There is some external arrival to the device's transmit buffer. The possible decisions are a) to serve some of the buffer content, b) to reorder a new battery after serving the maximum possible amount that it can, and c) to remain idle so that the battery charge can increase owing to diffusion process (possible in some commercially available battery). We look at both open-loop and closed-loop control of the system. For the closed-loop control, we view the problem in the framework of Markov Decision Processes and address finite and infinite horizon discounted costs as well as average cost minimization problems. *Without using any second order characteristics*, we obtain results that include i) optimality of bang-bang control, ii) the optimality of threshold based policies, iii) parameteric monotonicity of the threshold, and iv) uniqueness of the threshold. For the open-loop control setting we use recent advances in application of multimodular functions to establish optimality of bracket sequence based control.

Key-words: Markov Decision Process, Multimodularity

* V. S. Borkar is with School of Technology and Computer Science Tata Institute of Fundamental Research, Mumbai, 400 005, India. Email borkar@tifr.res.in. A. A. Kherani and B. J. Prabhu are with INRIA, 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France Email {alam, bprabhu}@sophia.inria.fr.
This work was supported by project no. 2900-IT-1 from the *Centre Franco-Indien pour la Promotion de la Recherche Avancee* (CEFIPRA).

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Contrôle de la file d'attente et de l'énergie d'un système mobile: approches en boucle fermée et en boucle ouverte

Résumé : Nous étudions un problème de décision en temps discret qui peut se poser pour un système mobile avec une source d'énergie limitée. Pendant chaque slot, il y a des paquets qui arrivent dans la file d'attente de cet système. Au début de chaque slot, l'système peut choisir parmi les décisions suivantes: a) envoyer quelques paquets, b) envoyer tous les paquets possible et commander une nouvelle batterie, et c) rester inactif, ce qui permet d'augmenter l'énergie résiduelle. Cette formulation mène à un compromis énergie/performance pour l'appraeil. Nous étudions le contrôle de ce système en boucle ouverte et en boucle fermée. Dans le cas d'un contrôle en boucle fermée, nous utilisons le cadre des processus de décision markoviens. L'objectif est de minimiser le coût actualisé sur un horizon de gestion fini et infini, et le coût moyen par unité de temps. Nous montrons, sans utiliser les caractéristiques de second ordre, i) l'optimalité de commande par tout ou rien, ii) l'optimalité de contrôle à seuil, iii) la monotonie paramétrique de ce seuil, et iv) l'unicité de ce seuil. Enfin, nous utilisons les progr̀es récents dans les applications des fonctions multimodulaires pour établir l'optimalité du contrôle basé sur des "bracket sequeunces".

Mots-clés : processus de décision markoviens, "Multimodularity"

1 Introduction

Wireless devices are constrained in their operational lifetime by finite energy batteries. Therefore, energy efficient design of protocols at different layers of the protocol stack for wireless networks has recently received significant attention, see, for example, [6]-[9]. Although the primary objective of a terminal is to transmit and receive data with minimal delay, this must be done with the added constraint of minimizing the transmission costs and increasing the operational lifetime of the terminal. Recently, in [10] the authors studied delay optimal packet scheduling policies subject to average transmit power constraint over a wireless channel with independent fading. In [3], the authors extended this model to include Markovian fading. Although in the above mentioned articles an average power constraint was imposed, an interesting feature of the battery was ignored. In [11], it was observed that a battery can regain some of its lost energy when left idle. This feature can enable a user to send more packets and increase the operational lifetime of the terminal if the terminal were to be left idle, thus providing incentive to remain idle even though the transmit buffer is not empty. However, this would add to the delay of the packets queued up in the buffer. This trade-off between energy and delay leads to a decision making problem formulation where the user has to decide whether to serve packets or leave the terminal idle in order to minimize certain costs.

In this study, we consider a discrete time system in which a user with a finite energy battery terminal has to decide whether to serve packets or to leave the system idle in each time slot. Further, the user can decide to replace the battery with a new one at an additional cost. We note that there are two variables (i.e., energy level of the battery and length of the transmit buffer) based on which a decision is to be made. We formulate the problem as a Markov decision process. We then derive the structural properties based on the directional derivative of the value function. We first consider a finite horizon problem and provide the structure of the optimal policy. We then extend this to the infinite horizon discounted cost problem, and finally consider the infinite horizon average cost minimization. We then consider the problem of making an optimal decision when the knowledge of the remaining energy and buffer occupancy are not known.

The outline of the paper is as follows. In Section 2 we formulate the problem of closed loop control. Sections 3 4.2 and 4.3 deal with finite horizon discounted cost, infinite horizon discounted cost and infinite horizon average cost respectively. Section 6 deals with open loop control.

2 Closed Loop Control

We first consider the optimal control problem where, at the beginning of each time slot (i.e., decision epoch), the device is aware of the current buffer occupancy and the energy level of the battery, and hence takes an action based on these two parameters. However, it does not have any knowledge of the amount of data that will be arriving to the transmit buffer in the current time slot. In this section, we shall formalize the problem statement mentioned in the Introduction.

2.1 Problem Formulation

Let x_n and p_n denote the buffer length and the remaining energy level, respectively, at the beginning of the n^{th} time slot. We assume x_n is infinitely divisible and $x_n \in [0, \infty), \forall n$, i.e., the buffer content is fluid and there is infinite buffer space. The remaining energy level, p_n, is assumed to be bounded above by M, i.e., $p_n \in [0, M], \forall n$. The state space, $\mathcal{C}$, is given by

$$\mathcal{C} = \{(x, p) : x \in [0, \infty), p \in [0, M]\}.$$

The cost function associated with the state (x, p) is denoted by

$$h(x) + g(p),$$

where $h(x)$ is an increasing function of x and $g(p)$ is a decreasing function of p. We note that we could also use a composite cost function $c(x, p)$ instead of $h(x) + g(p)$, and all the results in the paper would continue to hold. However, for simplicity, we use the above mentioned form. Let w_n denote the amount of fluid which arrives at the end of the n^{th} time slot. The random sequence $\{w_n, n \geq 0\}$, is assumed to be i.i.d with distribution $\mu(\cdot)$.

We assume that in every slot the battery is left idle, the residual battery energy increases from p to some amount $p + B(p) \geq p$. In the rest of the article, we will drop the dependence on p of $B(p)$ and use B to denote the function. We note that the case $B = 0$ corresponds to the other practical scenario where the battery does not gain its charge when left idle. In state (x, p), the user can take one of the following actions:

1. remain idle,
2. serve some amount $u \in [0, x \wedge p]$, or
3. serve $x \wedge p$ and order a new battery that has residual energy level M.

We denote the action space by $\mathcal{A}$, where $\mathcal{A} = \{1, 2, 3\}$. A cost of $r(p)$, where $r(\cdot)$ is a non-decreasing function of p, is incurred each time a battery is reordered in state (x, p). A policy π defines an action for each $(x, p) \in \mathcal{C}$. We look at optimal policies which minimize a given cost criterion. We consider the following three cost criteria.

- Finite horizon discounted cost

$$\sum_{k=0}^{N} \beta^k (h(x_k) + g(p_k)),$$

for some $N > 0$.

- Infinite horizon discounted cost

$$\sum_{k=0}^{\infty} \beta^k (h(x_k) + g(p_k)).$$

- Infinite horizon average cost

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} (h(x_k) + g(p_k)).$$

3 The Finite Horizon Case

For the finite horizon discounted cost case, the dynamic programming equations are

$$\begin{aligned} V_{k+1}(x,p) &= h(x) + g(p) + \beta \min \Big(E_w V_k((x-p)^+ + w, M) + r((p-x)^+), \\ & \min_{0\le u\le x\wedge p} E_w V_k(x+w-u, p-u), E_w V_k(x+w, p+B) \Big) \end{aligned} \tag{1}$$

where $V_k(x,p)$ is the optimal cost when the system state is (x,p) and there are k more decisions to be made before reaching the horizon. Since the decision epoch can be determined from V_k, we have used x, p and w instead of x_k, p_k and w_k, respectively. The operator E_w is the expection over the random variable w. This formulation allows us to consider the case where we don't buy a new battery when it is not completely drained out (i.e., $p > x$). An important property satisfied by the above formulation is that $\frac{\partial}{\partial p} r((p-x)^+) + \frac{\partial}{\partial x} r((p-x)^+) = 0$. In the following we will also be using a fixed cost c_3 for battery reordering instead of $r(\cdot)$.

For this section, we now assume that $\beta = 1$ with a note that all the results of this section are valid when we consider a discount factor $\beta < 1$. Let N denote the finite horizon. In this section we first provide simple conditions under which *bang-bang* control is optimal, i.e., the decision is to either serve the maximum possible quantity or to remain idle. These conditions are quite general and are satisfied by many natural candidates for the cost functions $h(x)$ and $g(p)$, in particular, for linear functions $h(x) = c_1 x$ and $g(p) = c_2 p$. We then provide structural result of the optimal policy, like parametric monotonicity of the threshold policy. We point out here that we obtain the results of this section *without using second order characteristics of the value functions*, like convexity or decreasing-differences property.

3.1 Optimality of Bang-Bang Control

Let $\nabla V(x,p) = \frac{\partial}{\partial y} V(y,q)|_{y=x} + \frac{\partial}{\partial q} V(y,q)|_{q=p}$ denote the directional derivate of $V(y,q)$ along the vector $(1,1)$ at (x,p). Assume that $h'(x) + g'(p)$ has same sign, say $\mathcal{S} \in \{+,-\}$, for all values of x and p. Under the above assumption, we have the following two lemmas.

Lemma 1 *For all values of $x, p, w,$ and k, $u^* = \underset{0 \le u \le x \wedge p}{argmin}\ E_w V_k(x+w-u, p-u)$ is given by*

$$u^* = \begin{cases} 0 & if \quad \nabla V_k(y,q) > 0, \forall (x,p), \\ x \wedge p & if \quad \nabla V_k(y,q) < 0, \forall (y,q). \end{cases}$$

Proof of Lemma 1:

$$\begin{aligned} E_w V_k(x+w-u, p-u) &= \int_w V_k(x+w-u, p-u)\mu(dw) \\ \nabla E_w V_k(x+w-u, p-u) &= \int_w \nabla V_k(x+w-u, p-u)\mu(dw). \end{aligned}$$

Since $\nabla V_k(x+w-u, p-u)$ has a sign independent of u, x, p, w, it follows that $\nabla E_w V_k(x+w-u, p-u)$ has the same sign as $\nabla V_k(x+w-u, p-u)$. Therefore, $E_w(V_k + w - u, p - u)$ is an increasing (decreasing) function of u when $\mathcal{S} = -(+)$. The lemma thus follows. •

Lemma 2 *$\nabla V_k(x,p)$ has sign $\mathcal{S}$ for all values of $k < N, x,$ and p.*

Proof of Lemma 2: The claim is true for $k = N-1$ as $V_{N-1} = h(x) + g(p)$. Now, assuming the claim to be true for some $k+1 \le N-1$, we get, using Lemma 1,

$$\begin{aligned} V_{k+1}(x,p) &= h(x) + g(p) + \min\Big(E_w V_k((x-p)^+ + w, M) + r((p-x)^+), \\ &\quad E_w V_k(x+w, p), E_w V_k(x+w-x\wedge p, p-x\wedge p), E_w V_k(x+w, p+B)\Big). \end{aligned}$$

From the hypothesis, the sign of $\nabla V_k(x,p)$ is $\mathcal{S}$ for all values of (x,p). As the expectation operator gives a convex combination, it follows from the hypothesis and envelop theorem [2, Lemma 3.3.1] that each of the terms under the minimum operator above will have the same property, i.e., their directional derivatives along vector $(1,1)$ will have sign $\mathcal{S}$. Recall that the term $r(\cdot)$ does not contribute to the directional derivative along the direction $(1,1)$. Now, since $h'(x) + g'(p)$ has sign $\mathcal{S}$, the directional derivative of $V_k(x,p)$ along $(1,1)$ will also have sign $\mathcal{S}$. •

Corollary 1 *For all values of $x, p, w,$ and $k < N$, $u^* = \underset{0 \le u \le x \wedge p}{argmin}\ E_w V_k(x+w-u, p-u)$, is $u^* = 0$ if $\mathcal{S} = -$ and $u^* = x \wedge p$ if $\mathcal{S} = +$.*

Proof: Follows from Lemma 1 and 2. •

Lemma 3 *$V_k(x,p)$ is decreasing in p for a fixed x and increasing in x for a fixed p.*

Proof: From the assumption, $g(p)$ (resp.$h(x)$) is decreasing (resp. increasing) in p (resp.x) and observe that $V_{N-1}(x,p) = h(x) + g(p)$. Proof follows by using induction approach as in previous proofs. •

Theorem 1 *Let $r(\cdot)$ be a constant, equal to c_3.*

1. *If $\mathcal{S} = -$ then optimal policy is to either a) serve maximum possible amount $(x \wedge p)$ and* ***reorder****, or b) remain idle.*

2. *If $\mathcal{S} = +$ then optimal policy is to either a) serve maximum possible amount $(x \wedge p)$ and* ***reorder****, or b) serve maximum possible amount $(x \wedge p)$ and* ***do not reorder****, or c) remain idle.*

Proof: For $\mathcal{S} = -$, the result follows from Corollary 1 and Lemma 3. For $\mathcal{S} = +$, the result follows from Corollary 1. •

Theorem 2 *If $\mathcal{S} = +$ and $p = M$ then optimal policy is to serve $x \wedge M$ and then decide to reorder or not.*

Proof: For a fixed value of w, we compare $V_{k+1}((x-M)^+ + w, (M-x)^+)$ with $V_{k+1}(x+w, M)$. Since $\mathcal{S} = +$, $V_{k+1}(x+w, M) \geq V_{k+1}(x+w-(x \wedge M), M-(x \wedge M)) = V_{k+1}((x-M)^+ + w, (M-x)^+)$. Since the above is true for each w, the proof follows. •
Remark When $p = M$ and $\mathcal{S} = +$, it is optimal not to leave idle.

We can actually strengthen Theorem 1 for the case $\mathcal{S} = +$ by using Lemma 3 in the following way:

Theorem 3 *Let $r(\cdot)$ be a constant, and let $B = 0$. For $\mathcal{S} = +$, the optimal policy is to either serve maximum possible amount $(x \wedge p)$ and* ***reorder****, or serve maximum possible amount $(x \wedge p)$ and* ***do not reorder****.*

Proof: For a fixed value of w, we compare $V_{k+1}((x-p)^+ + w, (p-x)^+)$ with $V_{k+1}(x+w, p)$. Since $\mathcal{S} = +$, $V_{k+1}(x+w, p) \geq V_{k+1}(x+w-(x \wedge p), p-(x \wedge p)) = V_{k+1}((x-p)^+ + w, (p-x)^+)$. Since the above is true for each w, the proof follows. •
Remark For $B = 0$ and $\mathcal{S} = +$, it is optimal at all decision epochs to serve $x \wedge p$. We are, therefore, left with the decision to **reorder** or **not to reorder**. However, for $B = 0$ and $\mathcal{S} = -$, we can not eliminate the decision to remain idle from the action space. This point will become clear later when we consider infinite horizon discounted cost problem.

The above series of results have systematically reduced the number of choices to be made. We also noted the optimality of bang-bang control policy, i.e, it was optimal to either serve the maximum possible amount or to serve nothing. Next we provide some structural results for the optimal policy. We study the cases $\mathcal{S} = -$ and $\mathcal{S} = +$ separately.

4 The Case of $\mathcal{S} = -$

In this section we consider the case $\mathcal{S} = -$ and obtain structural results for finite/infinite horizon discounted cost and then use *vanishing discount* approach to study the infinite horizon average cost case.

4.1 Structure of the Optimal Policy

We show that for each given value of p and k, there is a value $x_k^*(p)$ such that if $x < x_k^*(p)$ then optimal action is to remain idle. We also show that $x_k^*(p)$ is increasing function of p. We start with the following simple result:

Lemma 4 *The derivative* $\frac{d}{dp}V_k(x,p)$ *(resp.* $\frac{d}{dx}V_k(x,p)$*) is bounded above (resp. below) by* $\max_p \frac{d}{dp}g(p)$ *(resp.* $\min_x \frac{d}{dx}h(x)$*) for each* $k \leq N$*.*

Proof The statement is true for $k = N$. Since

$$V_k(x,p) = h(x) + g(p) + \min\{E_w V_{k+1}((x-p)^+ + w, M) + c_3, E_w V_{k+1}(x+w,(p+B)\wedge M)\},$$

and for each fixed value of w, $\frac{d}{dx}V_{k+1}((x-p)^+ + w, M)$ and $\frac{d}{dx}V_{k+1}(x+w,(p+B)\wedge M)$ are positive, we obtain that $\frac{d}{dx}V_k(x,p) \geq \min_p \frac{d}{dx}h(x)$. Using a similar argument, it follows that $\frac{d}{dp}V_k(x,p) \leq \max_p \frac{d}{dp}g(p)$. •

Remark If $\frac{d}{dp}g(p) < -1$ then $\frac{d}{dp}V_k(x,p) < -1$ for all values of $k \leq N$.

Since we are considering the case $\mathcal{S} = -$, from Theorem 1, we only have two actions to choose from, so that value iteration of Equation 1 becomes

$$V_{k+1}(x,p) = h(x) + g(p) + \beta \min\left\{E_w V_k((x-p)^+ + w, M) + c_3, E_w V_k(x+w,(p+B)\wedge M)\right\}.$$

We first define up crossover points and down crossover points as follows.

Definition 1 *For a given* k *and* p*,* $\overline{x}^*(k,p)$ *is called an up crossover point if there exist* $\epsilon > 0$ *and* $\delta > 0$ *such that the following three conditions are satisfied.*

1. $E_w V_k((x-p)^+ + w, M) + c_3 < E_w V_k(x+w,(p+B)\wedge M) \quad \forall x \in [\overline{x}^*(k,p) - \epsilon, \overline{x}^*(k,p)),$
2. $E_w V_k((x-p)^+ + w, M) + c_3 > E_w V_k(x+w,(p+B)\wedge M) \quad \forall x \in (\overline{x}^*(k,p), \overline{x}^*(k,p) + \delta],$ *and*
3. $E_w V_k((\overline{x}^*(k,p) - p)^+ + w, M) + c_3 = E_w V_k(\overline{x}^*(k,p) + w,(p+B)\wedge M).$

The down crossover points, $\underline{x}^*(k,p)$ *are defined by reversing the inequalities.*

For a fixed p and k, we note that $E_w V_{k+1}((x-p^+ + w, M) + c_3$ and $E_w V_{k+1}(x+w, p+B)$ are increasing functions of x. Therefore, we can define a sequence of up (resp. down) crossover points $\overline{x}_n^*(k,p)$ (resp. $\underline{x}_n^*(k,p)$).

Lemma 5 *For each step* k*, there is an* $x_k^*(p)$ *such that it is optimal to remain idle when* $x < x_k^*(p)$ *when the battery level is* p*.*

Proof We note that $x_k^*(p) = 0$ is always a possibility. •

Theorem 4 *If* $\mathcal{S} = -$*, i.e., the directional derivative* $\frac{d}{dx}h(x) + \frac{d}{dp}g(p) < 0$*, then* $x_k^*(p)$ *is increasing function of* p*.*

Proof It is enough to show that for each fixed value of w. the first crossover point of the curves $V_{k+1}((x-p)^+ + w, M) + c_3$ and $V_{k+1}(x+w, (p+B) \wedge M)$ is monotone increasing function of p. By first crossover point here we mean $x_w(p) = \inf\{x : V_{k+1}((x-p)^+ + w, M) + c_3 \leq V_{k+1}(x+w, (p+B) \wedge M)\}$. This implies that

$$\tfrac{d}{dx} V_{k+1}((x-p)^+ + w, M)|_{x=x_w(p)} < \tfrac{d}{dx} V_{k+1}(x+w, (p+B) \wedge M)|_{x=x_w(p)},$$
$$V_{k+1}((x_w(p)-p)^+ + w, M) + c_3 = V_{k+1}(x_w(p)+w, (p+B) \wedge M)$$

Now,

$$\begin{aligned}
V_{k+1}((x_w(p) - p - \delta)^+ + w, M) &+ c_3 - V_{k+1}(x_w(p) + w, (p+\delta+B) \wedge M) \\
&= V_{k+1}((x_w(p)-p)^+ + w, M) + c_3 - \delta \frac{d}{dx} V_{k+1}((x-p)^+ + w, M)|_{x=x_w(p)} \\
&\quad - V_{k+1}(x_w(p)+w, (p+B) \wedge M) - \delta \frac{d}{dp} V_{k+1}(x_w(p)+w, (p+B) \wedge M) \\
&= -\delta \frac{d}{dx} V_{k+1}((x-p)^+ + w, M)|_{x=x_w(p)} - \delta \frac{d}{dp} V_{k+1}(x_w(p)+w, (p+B) \wedge M) \\
&\geq -\delta \frac{d}{dx} V_{k+1}(x+w, (p+B) \wedge M)|_{x=x_w(p)} - \delta \frac{d}{dp} V_{k+1}(x_w(p)+w, (p+B) \wedge M) \\
&= -\delta \left[\frac{d}{dx} V_{k+1}(x+w, (p+B) \wedge M)|_{x=x_w(p)} + \frac{d}{dp} V_{k+1}(x_w(p)+w, (p+B) \wedge M) \right] > 0.
\end{aligned}$$

The proof is complete if we show that for $x < x_w(p)$, $V_{k+1}((x_w(p) - p - \delta)^+ + w, M) + c_3 - V_{k+1}(x_w(p) + w, (p+\delta+B) \wedge M) > 0$, because it will imply that an increase in p to $p+\delta$ does not change the ordering of the two terms for $x < x_w(p)$ so that the *first* crossover point is indeed increasing in p.

So, consider any point $x < x_w(p)$ so that $V_{k+1}((x-p)^+ + w, M) + c_3 > V_{k+1}(x+w, (p+B) \wedge M)$ and $V_{k+1}((x-p-\delta)^+ + w, M) + c_3 < V_{k+1}(x+w, (p+\delta+B) \wedge M)$. This implies that there is an $\epsilon < \delta$ so that

$$\tfrac{d}{dx} V_{k+1}((x-p-\epsilon)^+ + w, M)|_{x=x_w(p+\epsilon)} < \tfrac{d}{dx} V_{k+1}(x+w, (p+\epsilon+B) \wedge M)|_{x=x_w(p+\epsilon)},$$
$$V_{k+1}((x_w(p+\epsilon) - p - \epsilon)^+ + w, M) + c_3 = V_{k+1}(x_w(p+\epsilon) + w, (p+\epsilon+B) \wedge M).$$

Now we can use the method for $x_w(p)$ used above by taking $p = p + \epsilon$ and $\delta = \delta - \epsilon$ and arrive at a contradiction. •

Remark Note here that to obtain the parametric monotonicity we have not used convexity or decreasing differences property of the value function (which are in fact not true in our case). Similarly we get

Corollary 2 *If $\mathcal{X}(p)$ denotes the set of queue lengths x such that optimal decision is to remain idle when battery level is p, then $\mathcal{X}(p)$ is increasing in p in the sense that $\mathcal{X}(p) \subset \mathcal{X}(p+\delta)$, $\delta > 0$.*

4.2 Infinite Horizon Discounted Cost

Now we consider the case of $N = \infty$, i.e., infinite horizon problem. It is clear that all the properties obtained for the finite horizon problem are valid for this case also. For the finite horizon case, it was necessary to study the value function for all possible values of its argument, i.e., x and p. However, since we intend to study the average cost problem via the infinite horizon discounted cost case by using the *vanishing discount* approach, we will see that, for the case where $\mathcal{S} = -$, it is enough to study the value function $V(x,p)$ only at $p = M$. In this section we study the structure of the infinite horizon discounted problem value function $V(x, M)$ assuming $\mathcal{S} = -$. We start with considering the problem of minimizing the discounted cost

$$V(x,p) = \sum_{n=0}^{\infty} \beta^n (h(x_n) + g(p_n) + c_3 R_n)$$

where R_n is the indicator function of the event that the battery has been reordered in step n, thus incurring a cost c_3 and $x_0 = x, p_0 = p$. We know that $V(\cdot,\cdot)$ satisfies the Bellman equation

$$V(x,p) = h(x)+g(p)+\beta \min\{E_w V((x-p)^+ + w, M)+c_3, E_w V(x+w,(p+B)\wedge M), \min_{0\le u\le x\wedge p} E_w(x-u+w, p-u)\}.$$

Assuming that $\mathcal{S} = -$, we can use value iteration for the above problem to show that

$$V(x,p) = h(x) + g(p) + \beta \min\{E_w V((x-p)^+ + w, M) + c_3, E_w V(x+w, (p+B)\wedge M)\}.$$

Assume now that $h(x)$ is a concave differentiable increasing function so that $\sup_x \frac{dh(x)}{dx} = \frac{dh(x)}{dx}|_{x=0} \le \frac{(1-\beta)c_3}{M}$. Now we consider the value-iteration for $p = M$, i.e., assume $V_0(x,p) = h(x) + g(p)$ and apply the above minimization iteratively, generating a family of value functions $V_k(x,p),\ k \ge 0$ so that

$$V_{k+1}(x,p) = h(x) + g(p) + \beta \min\{E_w V_k((x-p)^+ + w, M) + c_3, E_w V({}_k x + w, (p+B)\wedge M)\}.$$

Since we assume $p = M$,

$$V_{k+1}(x,M) = h(x) + g(M) + \beta \min\{E_w V_k((x-M)^+ + w, M) + c_3, E_w V_k(x+w, M)\}.$$

Observe from the above expression that *once the battery energy reaches M, it stays there.* This simplifies the problem significantly as now, with initial battery level at M, the value function can be viewed as a function of only x, the buffer occupancy. We now prove that

Theorem 5 $\frac{dV_k(x,M)}{dx}|_{x=0} \le \frac{(1-\beta^{k+1})c_3}{M}$ *and $V_k(x,M)$ is concave in x for all $k \ge 0$.*

Proof By induction, observing that the statement is true for $k = 0$ by assumption. If $V_k(x,M)$ satisfies these two properties then for any x and w,

$$V_k(x+w, M) \le V_k((x-M)^+ + w, M) + c_3 \quad x \le M.$$

Since $V_k(x,M)$ is concave in x, and $V_k(M+w,M) \leq V_k(w,M)+c_3$, it follows that the above inequality is true even for $x > M$. Hence, $V_{k+1}(x,M) = h(x)+g(M)+\beta E_w V_k(x+w,M)$. Since $V_k(\cdot,M)$ is concave, it follows that $V_{k+1}(\cdot,M)$ is also concave and that $\frac{dV_{k+1}(x,M)}{dx}|_{x=0} \leq \frac{(1-\beta^{k+2})c_3}{M}$. •

Remark: The above result suggests that if $\sup_x \frac{dh(x)}{dx} = \frac{dh(x)}{dx}|_{x=0} \leq \frac{(1-\beta)c_3}{M}$ then it is optimal to remain idle *forever* whenever the battery is fully charged. This result, though seemingly counter-intuitive, can be explained by the fact that we are considering discounted cost which gives weightage to near future cost only. To be able to give more consideration to distant future, we require β very close to unity, in which case the condition of the theorem ($\sup_x \frac{dh(x)}{dx} = \frac{dh(x)}{dx}|_{x=0} \leq \frac{(1-\beta)c_3}{M}$) does not hold. This point will be clear when we consider the average cost optimization problem where one gives all weight to the distant future costs.

Remark: If $h(x)$ is not concave then we can only say that $\frac{dV_{k+1}(x,M)}{dx}|_{x=0} \leq \frac{(1-\beta)c_3}{M}$.

If we consider a linear form for the buffer cost, i.e., $h(x) = c_1 x$ then we have that

Theorem 6 *If $h(x) = c_1 x$ with $c_1 \leq \frac{(1-\beta)c_3}{M}$ then*

$$\frac{dV_k(x,M)}{dx}|_{x=0} = \frac{1-\beta^{k+1}}{1-\beta}c_1$$

so that $V_k(x,M)$ is linear in x for all $k \geq 0$.

Proof Trivial. •

Corollary 3 *If $h(x) = c_1 x$ with $c_1 \leq \frac{(1-\beta)c_3}{M}$ then*

$$V(x,M) = \frac{c_1 x + g(M)}{1-\beta}$$

Proof Follows from arguments similar to those used above. •

Theorem 7 *If $\mathcal{S} = -$, $h(x)$ is concave function such that and $\inf_x h'(x) = (1-\beta)(\frac{c_3}{M}+L)$ for $L > 0$ then there exists an N such that $V_N(x,M) > V_N(0,M)+c_3$ for some $x < M$.*

Proof By contradiction. Suppose that there exists no such N, i.e., for each k, $V_k(x,M) < V_k(0,M)+c_3$ for each $x < M$. Now,

$$V_1(x,M) = h(x)+g(M)+\beta\min\{E_w V_0(w,M)+c_3, E_w V_0(x+w,M)\}.$$

Since $V_0(x,M) = h(x)+g(M)$ is concave in x, it follows that if $V_0(x,M) < V_0((x-M)^+,M)+c_3$ for all $x < M$ then same is true for all x. Thus $V_1(x) = h(x)+g(M)+\beta E_w V_0(x+w,M)$ which again is concave in x and, by assumption, satisfies the property that $V_1(x,M) < V_1(0,M)+c_3$ for each $x < M$. Now, the derivative of $V_0(x,M)$ is at least $\sup_x h'(x) + \beta \inf_x h'(x) > sup_x h'(x)$ so that the slope of $V_1(x,M)$ is strictly more than that of $V_0(x,M)$. Proceeding in similar manner, we can show that slope of $V_k(x,M)$ for

$x < M$ is strictly increasing in k and is infact at least $\frac{1-\beta^k}{1-\beta}\inf_x h'(x) > (1-\beta^k)(\frac{c_3}{M}+L)$ as can be easily seen from the above. This contradicts the assumption that for each k, $V_k(x,M) < V_k(0,M)+c_3$ for each $x < M$. Hence the proof. •

Note that once the battery attains its maximum capacity, i.e., M, then if $\mathcal{S} = -$, the battery always remains fully charged. Hence if initially $p = M$ then we can consider $V_n(x,p)$ as a function of x alone (when considering infinite horizon problem). For this case we have the result of Theorem 8 which requires the following definition

Definition We say that a function $f(x)$ is M-convex if it is true that $f'(x) \leq f'(x+M)$ for all x.

Theorem 8 *If $\mathcal{S} = -$ and $h(x)$ is M-convex, increasing then $V_n(x,M)$ is continuous and M-convex, i.e.,*

$$\frac{d}{dx}V(x+M) \geq \frac{d}{dx}V(x), \quad x \geq 0.$$

Proof By induction. The above is true for $n = 1$ when $V_1(x) = h(x) + g(M)$. Assume above is true for all $k \leq n$. Then so that the required inequality is satisfied since

1. If $x < M$,

$$\begin{aligned} V_{n+1}(x+M) &= h(x+M) + g(M) + \beta \min\{E_w V_n(x+w) + c_3, E_w(x+M+w)\} \\ V_{n+1}(x) &= h(x) + g(M) + \beta \min\{E_w V_n(w) + c_3, E_w V_n(x+w)\}. \end{aligned}$$

2. If $x \geq M$,

$$\begin{aligned} V_{n+1}(x+M) &= h(x+M) + g(M) + \beta \min\{E_w V_n(x+w) + c_3, E_w(x+M+w)\} \\ V_{n+1}(x) &= h(x) + g(M) + \beta \min\{E_w V_n(x-M+w) + c_3, E_w V_n(x+w)\} \end{aligned}$$

Since each term in the minimum at x has derivative less than that of any of the terms under minimum operator at the point $x+M$, and since the derivative of $V_{n+1}(y)$ is atmost the derivative of any of the terms under the minimum operator plus $h'(y)$ and since $h(y)$ is assumed to be M-convex, $h'(x+M) \geq h'(x)$, the result follows. •

Remark The above result also implies that $V_n(x,M)$ is neither convex nor concave in general.

Now we assume, without loss of generality, that $\inf_x h'(x) = 1$ (this can always be done by appropriately scaling $g(p)$ and c_3); Note the equality above. We now use result from Lemma 4 to provide structure of the optimal policy.

Theorem 9 *If $\mathcal{S} = -$, $h(x)$ is M-convex increasing with $h'(0) = 1$ and $c_3 < M$ then there is a unique threshold T such that, for $p = M$, if $x \leq T$ then it is optimal to remain idle for the infinite horizon problem and it is optimal to serve $x \wedge M$ and reorder the battery otherwise.*

Proof Since Lemma 4 implies that the derivative of $V(x,M)$ is bounded below by unity, it follows that $V(M+x,M) \geq V(x,M)+M$. We thus have

$$\begin{aligned}
c_3 + E_w V((x-M)^+ + w)|_{x=M} &= c_3 + E_w V(w) \\
&= c_3 + \int_w V(w)\mu(dw) \\
E_w V(x+w)|_{x=M} &= E_w V(M+w) \\
&= \int_w V(M+w)\mu(dw) \geq M + \int_w V(w)\mu(dw).
\end{aligned}$$

Now, as $c_3 < M$, it follows from the above that $c_3 + E_w V((M-M)^+ + w) < E_w V(M+w)$ so that the optimal choice at $x = M$ is to serve all of M and reorder the battery.
Also, it is easy to see that at $x = 0$, $c_3 + E_w V((x-M)^+ + w) = c_3 + E_w V(w) > E_w V(w) = E_w V(x+w)$ so that at $x = 0$ it is optimal to leave the battery idle.

Now, assume that a point x is such that $c_3 + E_w V((x-M)^+ + w) < E_w V(x+w)$; we know that $x = M$ is such a point. Now, for any $y > 0$

$$\begin{aligned}
& c_3 + \int_w V((x+y-M)^+ + w)\mu(dw) \\
= \quad & c_3 + \int_w \left[V((x-M)^+ + w) + y\frac{d}{dx}V((x+y-M)^+ + w) + o(y) \right] \mu(dw) \\
<^{*1} \quad & \int_w \left[V(x+w) + y\frac{d}{dx}V((x+y-M)^+ + w) + o(y) \right] \mu(dw) \\
\leq^{*2} \quad & \int_w \left[V(x+w) + y\frac{d}{dx}V(x+y+w) + o(y) \right] \mu(dw) \\
= \quad & \int_w V(x+y+w)\mu(dw) + o(y)
\end{aligned}$$

where $*1$ follows from assumption on point x and $*2$ follows from Theorem 8. Thus, if it is optimal to serve $x \wedge M$ and reorder at x then the same is also optimal for all $y > x$, and in particular, for all $y > M$.

Since

- $E_w V(x+w)|_{x=0} < c_3 + E_w V((x-M)^+ + w)|_{x=0}$,
- $c_3 + E_w V((x-M)^+ + w)|_{x=0} = c_3 + E_w V(w)$, independent of x for $x < M$,
- $E_w V(x+w)|_{x=M} > c_3 + E_w V((x-M)^+ + w)|_{x=>}$,
- $E_w V(x+w)$ is continuous and its slope is strictly positive (at least unity),

it follows that there is a point $T < c_3$ such that $E_w V(x+w)|_{x=T} = c_3 + E_w V((x-M)^+ + w)|_{x=T}$ and that at any $T+\epsilon$ where $M-T > \epsilon > 0$, it holds that $E_w V(x+w)|_{x=T+\epsilon} < c_3 + E_w V((x-M)^+ + w)|_{x=T+\epsilon}$. Since we have already proved above that it is always optimal to serve and reorder battery whenever $x > M$, it follows that the same is true for all $x \geq T$ while for $x < T$, it is optimal to remain idle. •

Corollary 4 *Under conditions of Theorem 9, for each p, it is optimal to replace battery if* $x > T$.

Proof We use the already established parameteric monotonicity of the set $\mathcal{X}_p$ in Corollary 2. Now, consider any p and let $x_0(p) = \sup\{x : \text{it is optimal to remain idle for all } y \leq x\}$. Clearly, $x_0(p) \leq T$ (threshold of Theorem 9). because if otherwise, then the parameteric monotonicity obtained in Corollary 2 is violated. The proof thus follows. •

Theorem 7 can now be applied to the case where $h(x) = c_1 x$ and then Theorem 9 can be used to obtain more structural results for the case of $h(x) = c_1 x$. We thus have the following structure for the case of $\mathcal{S} = -$ and $h(x) = c_1 x$ (when starting from $p = M$):

1. if $c_1 \leq (1-\beta)\frac{c_3}{M}$, then it is optimal to always remain idle when using discount factor of β
2. otherwise, there is a $T < c_3$ such that it is optimal to remain idle for $x < T$ and replace battery for $x > T$.

where the first result is obtained from Theorem 5 and the second part is obtained as follows: since $c_1 > (1-\beta)\frac{c_3}{M}$, starting from $p = M$, in the value iteration we will ultimately get a stage at which $V(x, M) > V(0, M) + c_3$ for some $x < M$. Now, since the structure derived for $h(x)$ convex is valid here, Theorem 9 can be invoked and a similar proof yields the conclusion.

Let us now make the dependence of value function on β explicit and use $V_{k,\beta}(\cdot,\cdot)$ to represent the value function in k^{th} step.

Theorem 10 *If $c_3 < M$ then for each k and x, $V_{k,\beta}(x, M)$ is non-decreasing in β.*

Proof By Induction. The result is true for $k = 1$ because $V_{1,\beta}(x, M) = h(x) + g(M)$. Assume now that the above is true for some k then, for $\beta_1 < \beta_2$,

$$V_{k+1,\beta_1}(x, M) = h(x) + g(M) + \beta_1 \min\Big(E_w V_{k,\beta_1}((x-M)^+ + w, M) + c_3, E_w V_{k,\beta_1}(x+w, M)\Big)$$

and

$$V_{k+1,\beta_2}(x, M) = h(x) + g(M) + \beta_2 \min\Big(E_w V_{k,\beta_2}((x-M)^+ + w, M) + c_3, E_w V_{k,\beta_2}(x+w, M)\Big)$$

Now, we know that for $x > M$ the minimum is attained by reordering, hence

$$V_{k+1,\beta_1}(x, M) = h(x) + g(M) + \beta_1(E_w V_{k,\beta_2}((x-M) + w, M) + c_3)$$

and

$$V_{k+1,\beta_2}(x, M) = h(x) + g(M) + \beta_2(E_w V_{k,\beta_2}((x-M) + w, M) + c_3)$$

so that the assertion is true for $k+1$-step cost with $x > M$. Let x_k (resp. y_k) denote the threshold for k-th step when discount factor is β_1 (resp. β_2) Now assume $x < M$ so that there are the following possibilities:

1. $\max(x_k, y_k) < x$: For this case the assertion follows from arguments same as those used for $x > M$.

2. $x_k < x < y_k$: so that

$$V_{k+1,\beta_1}(x, M) = h(x) + g(M) + \beta_1(E_w V_{k,\beta_1}(w, M) + c_3)$$

and

$$V_{k+1,\beta_2}(x, M) = h(x) + g(M) + \beta_2 E_w V_{k,\beta_2}(x + w, M).$$

Since $E_w V_{k,\beta_1}(w, M) + c_3 \leq E_w V_{k,\beta_1}(x + w, M) \leq E_w V_{k,\beta_2}(x + w, M)$, the statement holds true for this case as well.

3. $x_k > x > y_k$: so that

$$V_{k+1,\beta_1}(x, M) = h(x) + g(M) + \beta_1 E_w V_{k,\beta_1}(x + w, M)$$

and

$$V_{k+1,\beta_2}(x, M) = h(x) + g(M) + \beta_2(E_w V_{k,\beta_2}(w, M) + c_3).$$

Since $E_w V_{k,\beta_1}(x+w, M) \leq E_w V_{k,\beta_1}(w, M)+c_3) \leq E_w V_{k,\beta_2}(w, M)+c_3)$, the statement again follows.

4. $\min(x_k, y_k) > x$: so that

$$V_{k+1,\beta_1}(x, M) = h(x) + g(M) + \beta_1 E_w V_{k,\beta_1}(x + w, M)$$

and

$$V_{k+1,\beta_2}(x, M) = h(x) + g(M) + \beta_2 E_w V_{k,\beta_2}(x + w, M).$$

The assertion is seen to be true for this case as well. •

Let $x_k(\beta)$ denote the unique threshold for k-step to go cost function when the discount factor is β.

Lemma 6 *The derivative* $\frac{d}{dx}V_{k,\beta}(x)$ *is non-decreasing function of* β.

Proof True for $k = 1$ where $V_{1,\beta}(x) = h(x)$ independent of β. Assume it is true for all $n \leq k$ then

$$\frac{d}{dx}V_{k+1,\beta}(x) = \begin{cases} h'(x) + \beta\frac{d}{dx}E_w V_{k,\beta}((x - M)^+ + w) & x > x_k(\beta) \\ h'(x) + \beta\frac{d}{dx}E_w V_{k,\beta}(x + w) & x < x_k(\beta) \end{cases}$$

so that the result follows for $k + 1$-step cost as well. •

Theorem 11 *The derivative* $\frac{d}{d\beta}x_k(\beta) \leq 0$.

Proof $E_w V_{k,\beta}(x+w)$ starts from $E_w V_{k,\beta}(w)$ at $x=0$ and reaches

$$E_w V_{k,\beta}(x_k(\beta)+w) = E_w V_{k,\beta}(w) + c_3$$

at $x = x_k(\beta)$ so that the total increase between 0 and $x_k(\beta)$ is exactly c_3. For an $\epsilon > 0$ such that $\beta + \epsilon < 1$, the function $E_w V_{k,\beta+\epsilon}(x+w)$ starts from $E_w V_{k,\beta+\epsilon}(w)$ at $x=0$ and reaches $E_w V_{k,\beta+\epsilon}(x_k(\beta)+w)$ at $x = x_k(\beta)$ Using Lemma 6, it follows that $E_w V_{k,\beta+\epsilon}(x_k(\beta)+w) \geq E_w V_{k,\beta+\epsilon}(w) + c_3$ because in between $x=0$ and $x = x_k(\beta)$, the slope of $E_w V_{k,\beta+\epsilon}(x+w)$ with respect to x is at least that of $E_w V_{k,\beta}(x+w)$. This implies that $x_k(\beta+\epsilon) \leq x_k(\beta)$. •

4.3 Average Cost

We now consider the problem of optimization of the infinite horizon average cost for the case $\mathcal{S} = -$. The approach is to use results from infinite horizon discounted cost optimization and then use the standard vanishing discount approach with $\beta \to 1$. It is clear that if average cost exist, it is independent of the initial state so that we can without loss of generality assume that $p_0 = M$. We saw that, for the discounted cost case with $\mathcal{S} = -$, once the level M is attained, it is retained throughout. Thus making the structural results obtained in Section 4.2 for this particular case very relevant to the analysis of average cost problem. We first state Conditions **W** that we will need in this section

W1 The state space is a locally compact space with a countable base.

W2 $R(x)$, the set of feasible actions in state x, is a compact subset of R (the action space), and the multifunction $x \to R(x)$ is upper semi continuous.

W3 The transition probability kernel of the controlled process is continuous in action with respect to weak convergencein the space of probability measures.

W4 The cost function is lower semi-continuous.

It is easily shown that the above conditions are satisfied in our problem.

A sufficient condition for existence of stationary average optimal policy, which can be obtained as limit of discounted cost optimal policies $f_\beta(x)$ is provided in [4]. In our problem $f_\beta : \mathcal{R} \to \{0,1\}$ where 0 means remaining idle and 1 means serving $x \wedge M$ and reordering. Define $w_\beta(x) = V_\beta(x) - \inf_x V_\beta(x)$:

Theorem 12 ([4] Theorem 3.8) *Suppose there exists a policy Ψ and an initial state x such that the average cost $V^\Psi(x) < \infty$. Let $\sup_{\beta<1} w_\beta(x) < \infty$ for all x and the Conditions **W** hold, then there exists a stationary policy f_1 which is average cost optimal and the optimal cost is independent of the initial state. Also f_1 is limit discount optimal in the sense that, for any x and given any sequence of discount factors converging to one, there exists a subsequence $\{\beta_m\}$ of discount factors and a sequence $x_m \to x$ such that $f_1(x) = lim_{m\to\infty} f_{\beta_m}(x_m)$.*

In order to apply above result we need to show:

1. Existence of policy Ψ: The policy of serving $x \wedge M$ in every slot yields a finite average cost.

2. $\sup_{\beta<1} w_\beta(x) < \infty$ for all x: For any β, since $V_\beta(x)$ is monotone increasing, it follows that $x^* := \underset{x}{\operatorname{argmin}} V_\beta(x) = 0$. We have

$$V_\beta(x^*) = h(0) + \beta E_w V_\beta(w)$$

Now, for any fixed $x_0 = x$, consider a policy that serves $x_j \wedge M$ and reorders till the first time the queue is empty (let us denote this time by a random variable Z). Then it is easily seen that

$$\begin{aligned} V_\beta(x) &\leq \sum_{n=\lceil \frac{x}{M} \rceil}^{\infty} [\sum_{j=0}^{n} \beta^j h(x + \sum_{i=1}^{j} w_i - (j+1)M) + \beta^n V_\beta(0)] P(Z=n) \\ &\leq \sum_{n=\lceil \frac{x}{M} \rceil}^{\infty} [\sum_{j=0}^{n} \beta^j h(x + \sum_{i=1}^{j} w_i - (j+1)M)] P(Z=n) + V_\beta(0) \\ \Rightarrow w_\beta(x) = V_\beta(x) - V_\beta(0) &\leq \sum_{n=\lceil \frac{x}{M} \rceil}^{\infty} [\sum_{j=0}^{n} h(x + \sum_{i=1}^{j} w_i - (j+1)M)] P(Z=n) \end{aligned}$$

the expression on the right hand side is independent of β and finite almost surely if $E[W] < M$. The required condition is thus verified.

Hence for average criteria the cost is independent of the initial state. So we can without loss of generality assume $p_0 = M$. So now making use of the results of Section 4.2 we obtain our main result

Theorem 13 *For the average cost optimization problem, if $\mathcal{S} = -$ and $h(x)$ is a convex function, then there exist a threshold based policy which gives the minimum cost.*

Proof We have seen that for the discounted cost case the optimal policy is threshold based. We also know that the threshold is a monotone function of the discount factor β. This, along with Theorem 12 implies the existence of a threshold based optimal policy for average cost case as well. •

5 $\mathcal{S} = +$ and $B \equiv 0$

We now consider the case where $\mathcal{S} = +$ and $B \equiv 0$. Our starting point is Theorem 3 which says that if $B \equiv 0$ then the optimal policy serves $x \wedge p$ and then decides whether to replace the battery or not. Thus, in this case the Bellman equations become

$$V_{k+1}(x,p) = h(x) + g(p) + \beta \min\{E_w V_k((x-p)^+ + w, M) + c_3, E_w V_k((x-p)^+ + w, (p-x)^+)\}.$$

The comparison in the minimization to be done here consists of two two terms $V_k((x-p)^+ + w, M) + c_3$ and $V_k((x-p)^+ + w, (p-x)^+)$. It is noted from these terms that the optimal policy should be a function only of $x - p$, i.e., the decision is same for all (x, p) for which $x - p$ is same. This structure helps us in accurately characterizing the optimal policy; this is done below.

5.1 Finite Horizon Discounted Cost

Consider the Bellman equation for the case $x \leq p$:

$$V_{k+1}(x,p) = h(x) + g(p) + \beta \min\{E_w V_k(w, M) + c_3, E_w V_k(w, p - x)\}.$$

Observe that the first term under the minimization operation is independent of x and p. Now, recall that we showed in Lemma 4 that $\frac{d}{dp} V_k(x,p)$ is bounded above by a negative quantity. Since $V_k(w, M) \leq V_k(w, M) + c_3$, it follows that for each w, there is a value $p_k^*(w)$ such that $V_k(w, l) > V_k(w, M) + c_3$ for all $l = p_k^*(w)$ and that $V_k(w, l) < V_k(w, M) + c_3$. Note here that it is possible that $p_k^*(w) = 0$ but what is important is that, owing to negative value of $\frac{d}{dp} V_k(x, p)$, the set $\{l : V_k(w, l) < V_k(w, M) + c_3\}$ is connected and hence has a smallest element that is $p_k^*(w)$. Note also that $p_k^*(w)$ *is independent of the state* (x, p). By taking expectation over w, we obtain

Theorem 14 *For a fixed $x < p$, there is a quantity p_k^* such that it is optimal to reorder battery when $x < p < x + p_k^*$ and it is optimal to not reorder battery when $x + p_k^* < p \leq M$.*

In order to derive the structure of the optimal policy for the case $x > p$, we again use Lemma 4. Consider the Bellman equation for the case $x > p$:

$$V_{k+1}(x,p) = h(x) + g(p) + \beta \min\{E_w V_k(x - p + w, M) + c_3, E_w V_k(x - p + w, 0)\}.$$

From Lemma 4, if $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$, then $\frac{d}{dp} V_k(x,p) \leq \frac{-c_3}{M}$ for all values of k thus implying that $E_w V_k(x - p + w, M) + c_3 \leq E_w V_k(x - p + w, 0)$. The condition $\sup_l \frac{d}{dl} g(l) < \frac{-c_3}{M}$ also implies that $p^*(w) > 0$ for all values of w, thus $p^* > 0$. Hence we have

Theorem 15 *If $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$ then it is optimal to reorder the battery after serving whenever $x > p - p_k^*$, for any k.*

This result, along with Theorem 14, gives complete structure of the optimal policy when $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$.

Now, note that if it turns out that $\frac{d}{dp} V_k(x,p) > -\frac{c_3}{M}$ for all values of x and p then $p^*(w) = 0$ for all w and that, if $x > p$,

$$V_k(x - p + w, M) + c_3 > V_k(x - p + w, 0) \quad \forall w$$

so that we get

Theorem 16 *If* $\inf_p \inf_l \frac{d}{dl} V_k(x,p) \geq \frac{-c_3}{M}$ *for all values of* k *then it is optimal to never reorder the battery.*

The results obtained here are very similar to those obtained for the case $\mathcal{S} = -$ in the sense that we get a threshold based policy where existence of a nontrivial threshold depends on the slope of the cost functions.

We now consider the infinite horizon discounted/average costs problems for this case.

For the average cost problem, we need equivalents of Lemma 6 and Theorem 11. Let $p_k^*(\beta)$ denote the unique threshold for k-step to go cost function when the discount factor is β.

Lemma 7 *The derivative* $\frac{d}{dp} V_{k,\beta}(x,p)$ *is non-increasing function of* β *for each* x*,* k *and* p*.*

Proof True for $k = 1$ where $V_{1,\beta}(x,p) = h(x) + g(p)$ independent of β. Assume it is true for all $n \leq k$ then

$$\frac{d}{dp} V_{k+1,\beta}(x,p) = \begin{cases} g'(p) + \beta \frac{d}{dp} E_w V_{k,\beta}((x-p)^+ + w, M) & p < x + p_k^*(\beta) \\ g'(p) + \beta \frac{d}{dp} E_w V_{k,\beta}(w, p-x) & p > x + p_k^*(\beta) \end{cases}$$

so that the result follows for $k+1$-step cost as well. •

Theorem 17 *The derivative* $\frac{d}{d\beta} p_k^*(\beta) \geq 0$*.*

Proof $V_{k,\beta}(w,p)$ starts from $V_{k,\beta}(w,0)$ at $x = 0$ and reaches

$$V_{k,\beta}(w, p_{k,\beta}^{(}w), 0) = V_{k,\beta}(w, M) + c_3$$

at $p = p_{k,\beta}^*(w)$ so that the total decrease between $p = p_{k,\beta}^*(w)$ and $p = M$ is exactly c_3. Since $\frac{d}{dp} V_k(x,p)$ is non-increasing function of β. It follows that $p_{k,\beta}^*(w) \leq p_{k,\beta+\epsilon}^*(w)$ for any $\epsilon \geq 0$, since the total decrease between $p = p_{k,\beta}^*(w)$ and $p = M$ is exactly c_3, independent of beta. •

We can now handle the infinite horizon discounted cost and averag cost problems in way similar to that done for $\mathcal{S} = -$. In particular, we get counterparts of Theorems 9 and 13.

Theorem 18 *If* $\mathcal{S} = +$*,* $g(p)$ *is decreasing with* $\inf_p g'(p) < -\frac{c_3}{M}$ *then there is a unique threshold* T *such that, for* $p = M$*, if* $x \leq T$ *then it is optimal to remain idle for the infinite horizon problem and it is optimal to serve* $x \wedge M$ *and replace the battery otherwise.*

Theorem 19 *For the average cost optimization problem, if* $\mathcal{S} = +$ *then there exist a threshold based policy which gives the minimum cost.*

6 Open Loop Control

The key result obtained for the average cost optimization problem was the existence of a threshold based policy for the case where $\mathcal{S} = -$. The problem with such an approach is that the threshold depends on the distribution of the arrival process so that the computation of the threshold becomes hard. We may also look at other *suboptimal* policies that can bee asily implementable. This can be done, for example, by restricting ourselves to policies that do not require the state information. Such policies need not be stationary. A possible decision problem now would be to find an optimal sequence of 0's and 1's where 0 corresponds to reordering a new battery and 1 means decision of remaining idle. We would like to have a bound on the long term average cost of reordering while at the same time would like to minimize the average buffer occupancy cost.

Let $\{a_n\}_{n\geq 1}$, $a_n \in \{0,\ 1\}$, be a sequence of controls so that $a_n = 1$ indicates decision of remaining idle at n^{th} decision instant and $a_n = 0$ implies serving $x_n \wedge M$ and reordering a new battery. We also require an upper bound on the rate at which the battery can be reordered. This can be done by letting

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq p$$

where p is chosen so that the system is also stable, i.e., the long term average of given service, $M \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (1 - a_n) \geq E[W]$, i.e., $M(1-p) \geq E[W]$.

We now have the problem of minimizing

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n(\mathbf{a})$$

subject to

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq p$$

where $\mathbf{a}$ is a control sequence and $x_n = (x_{n-1} - (1-a_n)M)^+ + w_n$ is the buffer occupancy at decision instant n. Without loss of generality we assume that $x_0 = 0$ so that $x_1 = w_1$.

Definition 2 ([5]) *A function $f:\ \{0,1\}^N \to \mathcal{R}$ is multimodular if*

$$f(a+v) + f(a+u) \geq f(a) + f(a+u+v)$$

for all $a \in \{0,1\}^N$ and all $v, u \in \mathcal{F}$, $v \neq u$ with

$$\mathcal{F} = \{-e_1, s_2, s_3, \ldots, s_N, e_N\}$$

where

$$\begin{array}{rcl} -e_1 & = & (-1 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0), \\ s_2 & = & (1 \quad -1 \quad 0 \quad \dots \quad 0 \quad 0), \\ s_3 & = & (0 \quad 1 \quad -1 \quad \dots \quad 0 \quad 0), \\ & & \vdots \\ s_N & = & (0 \quad 0 \quad 0 \quad \dots \quad 1 \quad -1), \\ e_N & = & (0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 1). \end{array}$$

Theorem 20 *The function $f_N(\mathbf{a}) := x_N(\mathbf{a})$ is multimodular for each N.*

Proof Fix a sample path $\{w_n\}_{1\leq n\leq N}$ of arrival process. For this sample path of arrival process, let $0(a)$ denote the set of decision instants n with $a_n = 0$ and $x_n \leq M$ so that the decision results in an empty queue and that $x_{n+1} = w_{n+1}$ (note that the set depends crucially on the sample path of arrival process). An important property is that if $j \notin 0(a)$ then $f_N(a) = f_N(a+s_j)$ and if $j \in 0(a)$ then $f_N(a) \geq f_N(a+s_j)$ (since $x_N(a) \geq x_N(a+s_j)$). We now have following possibilities:

- $v = s_j,\ 2 \leq j \leq N$: For this case we have following candidates for u:
  - $u = s_k, k > j$: (the case $k > j$ is similar): It is clear that if $j \notin 0(a)$ then $j \notin 0(a+u)$ hence $f_N(a+v) = f_N(a)$ and $f_N(a+u+v) = f(a+u)$ so that the required inequality is trivially satisfied for all values of k. Now, if $j \in 0(a)$ then we have that
    - $k \in 0(a+v)$: This case is studied considering two more possibilities:
      - $k \in 0(a)$: Then $f_N(a) = f_N(a+v)$; this is because then $x_k(a+v) = x_k(a)$. Since $f_N(a+u+v) \leq f_N(a+u)$, this case is checked.
      - $k \notin 0(a)$: Then $f_N(a) = f_N(a+u)$ and since $f_N(a+u+v) \leq f_N(a+v)$ this case is checked.
    - $k \notin 0(a+v)$: then $k \notin 0(a)$ because if $k \in 0(a)$ then $k \in 0(a+v)$ also. Then $f_N(a+u+v) = f_N(a+v)$ and since $f_N(a+u) = f_N(a)$ the relation is verified.
  - $u = e_N$: For this case it is clear that $f_N(a+u) \geq f_N(a)$ for any control sequence a. Also, for this case $f_N(a+v) - f_N(a+v+u) \leq f_N(a) - f_N(a+u)$ (this is because $x_N(a) - x_N(a+e_N)$ is exactly $x_{N-1}(a) + w_N$ for any control a and since in our case $x_{N-1}(a+v) \leq x_{N-1}(a)$). This is the required inequality itself.
  - $u = -e_1$: Since we are starting with an empty system ($x_0 = 0$), this case is also trivially satisfied as it does not matter whether $a_0 = 0$ or 1. Since this also applies for the case where $v = e_N$ and $u = e_N$, multimodularity of the function $f_N(a)$ is established. •

We now state a key result from [5], but first we need the following definition

Definition 3 *Let p and θ be two positive reals. The bracket sequence $\{a_k^p(\theta)\}$ with rate p and initial phase θ is defined as $a_k^p(\theta) = \lfloor kp+\theta \rfloor - \lfloor (k-1)p+\theta \rfloor$.*

Theorem 21 (Theorem 6, [5]) *Under conditions*

<1> f_k *is multimodular,*

<2> $f_k(a_1, \ldots, a_k) \geq f_{k-1}(a_2, \ldots, a_k),\ \forall k > 1,$ *and*

<3> For any sequence $\{a_k\} \exists$ *a sequence* $\{b_k\}$ *such that* $\forall k, m$ *with* $k > m$, $f_k(b_1, \ldots, b_{k-m}, a_1, \ldots, a_m) \leq f_m(a_1, \ldots, a_m)$,

and given some $p \in [0, 1]$, *if the functions* $f_k(a_1, \ldots, a_k)$ *are increasing in all* a_i, *then the bracket sequence* $a^p(\theta)$ *for any* $\theta \in [0, 1)$, *minimizes the average cost*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(a)$$

over all sequences that satisfy the constraint

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq p.$$

Now we make the following assumption: The maximum amount of arrival in a slot is bounded by M. Under this assumption, all the conditions in the above theorem are satisfied (we can take $b_n \equiv 0$ since $W < M$). We have thus established the optimality of bracket sequences (of rate p) for open loop control of the system under consideration.

7 Conclusion

We considered jointly optimal scheduling and power control of a wireless device and formulated it as a Markov decision process problem. We considered the cases of optimizing finite and infinite horizon discounted costs as well as that of infinite horizon average cost. The problem becomes hard as the underlying state space is two-dimensional and important second order properties like convexity or increasing/decreasing differences do not hold. We established the optimality of bang-bang policy which is threshold based. We also studied the behaviour of this threshold and obtained parameteric monotonicity results. We then considered the problem of open-loop control of the system where the decision maker does not have knowledge of the system state. For this case we proved that using a bracket sequence based policy results in optimal performance.

There are still some issues that need to be addressed in order to complete the study of the problem. Some of these are

1. In the study of open loop control, we assumed that the amount of arrival in any slot is bounded by M. This may not be valid in general. Hence removing this assumption is also important.

2. Study of open loop control for the case where $\mathcal{S} = +$ so that the decisions to be made after serving maximum possible is whether to reorder battery or not.

Before concluding, we remark that the finite horizon model considered in this work is that of a very general interacting system where a change in one state results in corresponding change in the other dimension. This is, for example, the case in tandem queues where a departure from first queue results in an arrival to the second queue. The use of the sign of directional derivative $\mathcal{S}$ to establish optimality of bang-bang control for such systems appear to be natural choice. One study of control of tandem queue that appears to be very similar to our problem is that of [1] (they consider linear cost functions). However, they assume that there is always a dummy customer present in the second queue so that the term corresponding to our $(x - M)^+$ term does not show up in their case; this is also the reason that they are able to show convexity of the value function.

Also, the analysis of the infinite horizon case (with $p = M$) is applicable to control of service of a discrete time queue where amount of service in a slot is deterministic (M) and there is a fixed cost c_3 on each service and in any slot the server has to decide whether to serve or not.

References

[1] Z. Rosberg, P. P. Varaiya and J. C. Walrand. Optimal Control of Service in Tandem Queues IEEE Transactions on Automatic Control, vol. AC-27, no. 3, June 1982.

[2] D. P. Bertsekas. Dynamic Programming and Optimal Control. Athena Scientific.

[3] G. S. Rajadhyaksha and V. S. Borkar. Transmission Rate Control Over Randomly Varying Channels. Probability in Engineering and Informational Sciences, Cambridge Univ. Press, 2003.

[4] Manfred Schal. Average Optimality in Dynamic Programming with General State Space. Mathematics of Operations Research, vol. 18, 1993.

[5] E. Altman, B. Gaujal and A. Hordijk. Discrete-event control of stochastic networks: Multimodularity and Regularity Springer Verlag, Series: Lecture Notes in Mathematics 2003.

[6] S. Cui, A. J. Goldsmith and A. Bahai. Energy-constrained modulation optimization. *To appear in IEEE Trans. on Wireless Communications,*, 2004.

[7] W. Ye, J. Heidemann and D. Estrin. An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In *Proc. of IEEE INFOCOM*, pages 1567–1576, 2002.

[8] Michele Zorzi and Ramesh R. Rao. Energy Efficiency of TCP in a Local Wireless Environment. *Mob. Netw. Appl.*, 6(3):265–278, 2001.

[9] C. E. Price, K. M. Sivalingam, P. Agarwal and J.-C. Chen. "A Survey of Energy Efficient Network Protocols for Wireless and Mobile Networks. *ACM/Baltzer Journal on Wireless Networks*, 7(4):343–358, 2001.

[10] Munish Goyal, Anurag Kumar and Vinod Sharma. Power Constrained and Delay Optimal Policies for Scheduling Transmissions over a Fading Channel. In *Proc. of the IEEE INFOCOM*, 2003.

[11] T. F. Fuller, M. Doyle and J. S. Newman. Relaxation phenomena in lithium-ion insertion cells. *J. Electrochem. Soc.*, 141:982–990, 1994.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399